N72-32587

NASA CONTRACTOR REPORT

NASA CR-2119

NASA CR-2119

DIFFERENTIAL CORRECTION SCHEMES IN NONLINEAR REGRESSION

by Henry P. Decell, Jr., and F. M. Speed

Prepared by
UNIVERSITY OF HOUSTON
Houston, Texas
for Manned Spacecraft Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • SEPTEMBER 1972

1. Report No.
NASA CR-2119

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle

DIFFERENTIAL CORRECTION SCHEMES IN NONLINEAR REGRESSION

5. Report Date
September 1972

6. Performing Organization Code

7. Author(s)
Henry P. Decell, Jr., and F. M. Speed

8. Performing Organization Report No.
MSC S-352

9. Performing Organization Name and Address
Manned Spacecraft Center
Houston, Texas 77058

10. Work Unit No.

11. Contract or Grant No.
NAS 9-12777

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Washington, DC 20546

13. Type of Report and Period Covered
Contractor Report

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This paper briefly reviews and improves upon classical iterative methods in nonlinear regression. This is accomplished by discussion of the geometrical and theoretical motivation for introducing modifications using generalized matrix inversion. Examples having inherent pitfalls are presented and compared in terms of results obtained using classical and modified techniques. The modification is shown to be useful alone or in conjunction with other modifications appearing in the literature.

17. Key Words (Suggested by Author(s))
- Nonlinear Regression
- Differential Correction Schemes
- Generalized inverse
- Gauss-Newton Method
- Least Squares

18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
None	None	16	$3.00

* For sale by the National Technical Information Service, Springfield, Virginia 22151

DIFFERENTIAL CORRECTION SCHEMES IN NONLINEAR REGRESSION

Henry P. Decell, Jr.

University of Houston, Houston, Texas

F. M. Speed

Texas A & I University

Abstract

This paper briefly reviews and improves upon classical iterative methods in nonlinear regression. This is accomplished by discussion of the geometrical and theoretical motivation for introducing modifications using generalized matrix inversion, other than but in the same general vein as those discussed by Fletcher [6]. Examples having inherent pitfalls described in [8], [12] and others are presented and compared in terms of results obtained using classical and modified techniques. The modification is shown to be useful alone or in conjunction with other modifications appearing in the literature.

Introduction

Following for convenience the notation of [8], let y_t denote a set of n responces of the form

$$y_t = f_t(\theta) + e_t \ , \qquad t = 1,\ldots,n$$

where the response function $f_t(\theta)$ is a known function of t and an undetermined vector $\theta = (\theta_1,\ldots,\theta_p)$. We will call the vector $\hat{\theta}$ a least-squares estimate (given the n responses) of θ provided $\hat{\theta}$ minimizes

$$Q(\theta) = \sum_{t=1}^{n} (y_t - f_t(\theta))^2 \ .$$

The vectors are defined

$$Q'(\theta) = \frac{\partial(Q(\theta))}{\partial\theta_i}$$

$$R(\theta) = (y_t - f_t(\theta))$$

and the matrices

$$f'(\theta) = \left(\frac{\partial(f_t(\theta))}{\partial\theta_i} \right)^T$$

$$Q''(\theta) = \frac{\partial\left(\frac{\partial Q(\theta)}{\partial\theta_j}\right)}{\partial\theta_i} \ .$$

Three of the most common differential correction schemes for estimating the parameter vector $\hat{\theta}$ are the steepest descent method, the quadratic approximation, and the Gauss-Newton method, with corrections respectively given by

$$\Delta\theta = -\alpha Q'(\theta) , \qquad \alpha > 0$$

$$\Delta\theta = -(Q''(\theta))^{-1}Q'(\theta)$$

$$\Delta\theta = -1/2(f'(\theta)^T f'(\theta))^{-1}Q'(\theta) .$$

These methods have their advantages and disadvantages. Of the three, the Gauss-Newton method is probably most popular.

The authors of [8] present a modification of a classical method and state that "The step $\Delta\theta$ will in general be distinct in both length and direction for each of the three methods." This is not necessarily the case from a computational point of view since the matrices to be inverted may be, for all practical computational purposes, singular; yet the system of equations may have infinitely many solutions. For example, the Gauss-Newton correction requires the solution of the equation

$$f'(\theta)^T f'(\theta)\Delta\theta = f'(\theta)^T R(\theta)$$

since

$$-1/2Q'(\theta) = f'(\theta)^T R(\theta) .$$

It is known that any equation of this form (i.e., of the form $A^TAx = A^Tz$, the normal equations of the least-squares problem: minimize $(Ax-z)^T(Ax-z)$ given A and z) always has at least one solution and perhaps infinitely many. We will try to point out the significance and consequences of these solutions in terms of their relationship to differential correction schemes.

The Generalized Inverse

A few basic concepts regarding generalized inverses important to the development follow.

Theorem 1. The four equations $AXA = A$, $XAX = X$, $(AX)^* = AX$, and $(XA)^* = XA$ have a unique solution X for each complex $m \times n$ matrix A. This solution X is called the generalized inverse of A and is denoted by $X = A^+$.

This theorem is due to Penrose [10] and is equivalent to the apparently more geometric characterization of the generalized inverse of A which follows.

Theorem 2. The generalized inverse A^+ of A is the unique solution of the equations

$$AX = P_{R(A)}$$

$$XA = P_{R(X)}$$

where $P_{R(A)}$ and $P_{R(X)}$, respectively, denote the perpendicular projection operators on the range spaces (column spaces) of A and X.

In any case, it is easy to see that if A is square and non-singular, then A^+ is the ordinary inverse of A. Much work has been done recently in the area of generalized matrix inversion, including theoretical developments and computational techniques, rendering it a very useful tool in matrix theory and applications. A rather exhaustive bibliography concerning applications of generalized inverses can be found in [2], [3], and [13]. We will not develop the details of the basic concepts, but rather state an important theorem regarding the solution of matrix equations in general.

Theorem 3. The matrix equation $AXB = C$ has a solution X if and only if $AA^+CB^+B = C$, in which case all solutions are given by

$$X = A^+CB^+ + S - A^+ASBB^+$$

where S is an arbitrary matrix having the dimensions of X.

The Equation $A^TAx = A^Tz$

As stated earlier, the Gauss-Newton method involves the solution of an equation of this type at each iteration. The following corollary to Theorem 3 will give some insight to a possible course of action one could take at those times during the iteration process when the matrix $f'(\theta)^T f'(\theta)$ (or perhaps even a matrix such as $Q''(\theta)$ in another method

requiring inversion for the calculation of the correction $\Delta\theta$) is actually or nearly singular. For the purpose of this paper, we will describe how generalized inversion can be useful in iterative techniques requiring the solution of equations of the form $A^TAx = A^Tz$.

<u>Corollary 1</u>. If A is any $m\times n$ matrix and z is any $m\times 1$ vector, then the equation $A^TAx = A^Tz$ has at least one solution and all solutions are given by

$$X = A^+z + (I - A^+A)y$$

where y is arbitrary having the dimensions of x.

The proof of Corollary 1 is an immediate consequence of Theorem 3 and fact that $(A^TA)^+A^T = A^+$ [10].

<u>Corollary 2</u>. Among the solutions of $A^TAx = A^Tz$, the solution $x = A^+z$ has the smallest Euclidean norm (henceforth "norm" will be denoted $||\cdot||$).

The proof of Corollary 2 follows from the facts that $I - A^+A$ is the orthogonal projection operator on the orthogonal compliment of the range space of A^+ and hence that A^+z and $(I - A^+A)y$ are orthogonal for every y. In fact,

$$\begin{aligned} ||A^+z + (I - A^+A)y||^2 &= ||A^+z||^2 + ||(I - A^+A)y||^2 \\ &\geq ||A^+z||^2 . \end{aligned}$$

The significance of Corollary 1 is that there may be infinitely many possible corrections $\Delta\theta$ satisfying an equation defining a differential correction scheme in the presence of a singular or, in the computational sense, nearly singular coefficient matrix. There is a tendency to disregard or remain unaware of these solutions and, with the inability to invert the coefficient matrix, to look for new or modified techniques such as those found in [1], [5], [8], [9], and [12]. For example, in [7] Jennrich and Sampson modify the coefficient matrix by selected rows and columns. In [8], Marquardt changes the diagonal of the coefficient matrix. It has been our experience that these solutions should be given careful attention in the case of what will hereafter be called an apparent (i.e., actual or computational) singularity.

Fletcher [6] points out that in the generalized least-squares (Gauss-Newton) or Newton methods "... A most important property of the generalized inverse formulation is that in all circumstances (i.e., full rank or not), even when the generalized least-squares method would fail, the directions of search generated are downhill and so an imporvement can always be made to the sum of squares (assuming the approximation is not already a stationary point)." In this connection, the significance of Corollary 2 is that there is a reasonable way to choose a correction $\Delta\theta$ satisfying the defining equations of the scheme whenever an apparent singularity occurs. We propose to choose the minimum Euclidean norm correction $A^{+}z$ (i.e., the correction of shortest length consistent with the correction equation). It has been our experience that in nonlinear

equations other solutions can result in failure of convergence.

The suggested correction certainly depends upon the algorithm used to calculate A^+ and the actual computational way in which the algorithm establishes that A is not of full rank (i.e., A^TA singular). Of course, this is intimately connected with near-zero tests in the algorithm, sensitivity to dependent columns or rows, conditioning, and so forth. We should further point out that, for a general differential correction scheme of the form $M(\theta)\Delta\theta = z(\theta)$, the choice of the correction should be $\Delta\theta = M(\theta)^+z(\theta)$ if there is at least one solution for $\Delta\theta$. Of course, according to Theorem 3 there will be at least one and possibly infinitely many solutions $\Delta\theta$ if and only if $M(\theta)M(\theta)^+z(\theta) = z(\theta)$. Moreover, if there is one and only one solution, then that solution is indeed given by $\Delta\theta = M(\theta)^+z(\theta)$.

For example, in the Gauss-Newton method, $M(\theta) = f'(\theta)^Tf'(\theta)$ and $z(\theta) = f'(\theta)^TR(\theta)$ so that $\Delta\theta = M(\theta)^+z(\theta) = (f'(\theta)^Tf'(\theta))^+f'(\theta)^TR(\theta) = f'(\theta)^+R(\theta)$. Even if $M(\theta)$ is nonsingular, then $(f'(\theta)^Tf'(\theta))^+ = f'(\theta)^Tf'(\theta))^{-1}$, and either form of $\Delta\theta$ may be used in calculations:

$$\Delta\theta = (f'(\theta)^Tf'(\theta))^{-1}f'(\theta)^TR(\theta) = f'(\theta)^+R(\theta) .$$

In other words if $M(\theta)$ is square and computationally nonsingular, the classical correction is, in fact, the minimum norm correction. We will not discuss the comparative aspects of computing $\Delta\theta$ in a correction scheme such as the Gauss-Newton method by one or the other of the

theoretically equivalent formulas:

$$(1)\quad \Delta\theta = (f'(\theta)^T f'(\theta))^+ f'(\theta)^T R(\theta)$$

$$(2)\quad \Delta\theta = f'(\theta)^+ R(\theta)$$

Calculations in our examples use (2).

We have had unusual success with this technique in many practical problems too numerous to mention here. In many cases, one definite advantage seems to be the ability to continue making corrections of reasonable length and perhaps, as in the Gauss-Newton case, reasonable direction through regions in which the coefficient matrix $M(\theta)$ behaves badly. We do not propose this technique as a cure-all but rather that it should be included among other useful techniques in nonlinear regression. A few examples having known pitfalls will be presented in the next section.

Examples.

In the following examples, the residual sum of squares $Q(\theta)$ will be presented in tables by iteration number. The values of $Q(\theta)$ for the methods cited will be those values tabulated in the references cited. Some authors divide $Q(\theta)$ by the degrees of freedom. For clarity and easy comparison we indicate this division in the tables when necessary. Finally, the residual sum of squares given by the method of this paper (minimum norm correction) will be noted MN; $Q(\theta)$.

Results of the method of this paper compared with those of the Modified Davidon Method (MDM) used in [12] to find the parameters of an exponential model discussed by Hartley in [7] are given in Table 1.

Table 1

Exponential Model (Hartley)

Iteration	MN; $Q(\theta)$	MDM; $Q(\theta)$
0	27376	27376
1	14586	20127
2	13779	15412
3	13408	13552
4	13394	13485
5	13390	13449
6		13425
7		13394
8		13393
9		13390

A second exponential model given by the authors of [8] points out a failure of Hartley's method [7] due to a singular partial derivative matrix. In [8] a stepwise regression scheme (SR) is successfully utilized for this example. The results of the (SR) scheme compared with those of the method of this paper are given in Table 2.

Table 2

Exponential Model - Singular Partials

Iteration	MN; $Q(\theta)/8$	SR; $Q(\theta)/8$
0	521.41	521.41
1	429.84	429.84
2	39.11	88.15
3	15.765	83.74
4	15.545	*
10		21.33
30		15.545

*The value of SR: $Q(\theta)$ was not tabulated in [8] for this iteration.

Another six-parameter exponential model having inherent singularity problems is presented in [12] using the Modified Davidon Method (MDM). A comparison of the results using the technique of this paper is given

in Table 3.

Table 3

Six Parameter Exponential Model - Singular Partials

Iteration	MN; $Q(\theta)$	MDM; $Q(\theta)$
0	21.38	21.38
10	.873	2.39
20	.792	1.99
30	.396	1.77
40		1.59
50		1.41
60		.90
70		.41
80		.407

Concluding Remarks

We have taken the liberty to exclude a reproduction of the detailed description of our example models. These models are thoroughly treated in [7], [8] and [12]. The tables give some indication of rates of convergence and a comparison of residuals only. We do not wish to leave the impression

that iteration counts are comparable. For example, one Gauss-Newton iteration could have been equivalent to p conjugate direction steps for the matrix inversion employing the Davidon method.

Acknowledgments

The authors wish to express thanks to F. C. Delaney [4] for providing the detailed calculations presented in the tables. We would also like to express our thanks to G. E. Taylor [12] for making available to us the data used in his computations for MDM $Q(\theta)$ in Table 3.

REFERENCES

1. W. C. Davidon, Variable metric method for minimization, A.E.C. Research and Development Report ANL-5990 (Rev.), (1959).

2. H. P. Decell, An alternate form of the generalized inverse of an arbitrary complex matrix, SIAM Rev., (3) 7 (1965), 356-358.

3. ____________ , An application of generalized matrix inversion to sequential least squares parameter estimation, NASA TN D-2830, (1965).

4. F. C. Delaney, Generalized inverse calculation subroutine, (GINV2), Catalog 125, Lockheed Electronics Col, Houston, Texas, (1968).

5. R. Fletcher and J. J. D. Powell, A rapidly convergent descent method for minimization, The Computer Journal, 6 (1963), 163.

6. ____________ , Generalized inverse methods for the best least squares solutions of systems of nonlinear equations, The Computer Journal, (1968), 392-399.

7. H. O. Hartley, The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares, Technometrics, 3 (1961), 269-280.

8. R. I. Jennrich and P. F. Sampson, Application of stepwise regression to nonlinear estimation, Technometrics, 10 (1968), 63-72.

9. D. W. Marquardt, An algorithm for the estimation of nonlinear parameters, Society for Industrial and Applied Mathematics Journal, 11 (1963), 431-441.

10. R. Penrose, A generalized inverse for matrices, Proc. Camb. Philos. Soc., 51 (1955), 406-413.

11. J. B. Rosen, The gradient projection method for nonlinear programming, Part I. Linear Constraints, Society for Industrial and Applied Mathematics Journal, 8 (1960), 181-217.

12. P. Vitale and G. Taylor, A note on the application of Davidon's method to nonlinear regression problems, Technometrics, 10 (1968), 843-849.

13. Thomas L. Boullion and Patrick L. Odell, Proceedings of the symposium on theory and application of generalized inverses of matrices, Mathematics Series No. 4, Texas Technological College, Lubbock, Texas (1968).

14. G. Taylor, Data for example stated in [11], Private Communication.

NASA-Langley, 1972 —— 19 CR-2119

NATIONAL AERONAUTICS AND SPACE ADMISTRATION
WASHINGTON, D.C. 20546

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE $300

FIRST CLASS MAIL

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

U.S.MAIL

NASA 451

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION OFFICE

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546